Neutrosophic Decision Boundary Geometry (NDBG):

Connecting Smarandache Geometries and Machine Learning Interpretability

PREPRINT · ResearchGate · Not peer-reviewed

· **Maikel Leyva-Vázquez[1] Florentin Smarandache[2]**

[2] *Universidad de Guayaquil; Universidad Bolivariana del Ecuador · maikel.leyvav@ug.edu.ec*

[1] *University of New Mexico, Gallup, NM, USA · smarand@unm.edu*

Abstract

We introduce Neutrosophic Decision Boundary Geometry (NDBG), a framework establishing that the feature space of any differentiable classifier is a Smarandache Geometry where the classification axiom is S-denied in a computable, graduated manner. Three geometric invariants of the decision boundary $B = f^{-1}(0.5)$ define the Neutrosophic Geometric Field $\Phi(x) = (T(x), I(x), F(x))$: boundary distance (Truth T), Hessian curvature operationalizing the S-negation degree (Indeterminacy I), and adversarial vulnerability radius (Falsity F). The confidence score $C(x) = \max(0, T-I-F)$ is derived from the SVNS score function without free parameters, yielding the abstention rule $T \leq I+F$. Four theorems prove geometric independence of T, I, F; Smarandache zone decomposition $M = \Omega_E \cup \Omega_N \cup \Omega_A$; hyper-neutrosophic boundary states $T+I+F > 1$; and divergence of d_N from d_{Eucl} in NeutroGeometry zones. Geometric validation on Make Moons confirms all predictions. NDBG provides the first computable operationalization of the S-negation degree and unifies neutrosophic logic, neutrosophic sets, and Smarandache geometry in a single machine learning interpretability framework.

Keywords: neutrosophic geometry; Smarandache geometry; S-denied axiom; decision boundary; SVNS score; machine learning interpretability; adversarial robustness

1. Motivation

Classical uncertainty quantification in machine learning reduces prediction reliability to a scalar — probability P(y|x) — collapsing three epistemically distinct phenomena into one number. A query near the decision boundary $B = f^{-1}(0.5)$ may be uncertain because it is geometrically close to B (low separation), because the boundary is locally curved and unstable at that point (near-bifurcation behavior), or because a small adversarial perturbation suffices to cross B (topological fragility). Each phenomenon has a different actionable interpretation; conflating them loses information critical for model diagnosis and safe deployment.

Neutrosophic logic [1, 2] naturally represents tripartite uncertainty via independent components T, I, F. Existing applications assign T, I, F from training data proximity — not from the geometric structure of the learned model itself. NDBG derives all three from geometric invariants of B, yielding a post-hoc, model-centric framework applicable to any pretrained differentiable classifier.

2. The NDBG Framework

2.1 Neutrosophic Geometric Field

Definition 1. Let $f: \mathbb{R}^n \to [0,1]$ be a differentiable classifier. The decision boundary $B = f^{-1}(0.5)$ is a smooth $(n-1)$-dimensional submanifold. The Neutrosophic Geometric Field is $\Phi: \mathbb{R}^n \to [0,1]^3$:

$$T(x) = 2 \cdot |f(x) - 0.5| \quad \text{[distance to B — Truth]}$$

$$I(x) = \min(1, \|H_f(x)\|_F / \kappa) \quad \text{[Hessian curvature — Indeterminacy]}$$

$$F(x) = 1 - \min(1, \|\delta^*(x)\| / \varepsilon) \quad \text{[adversarial radius — Falsity]}$$

$H_f(x)$ diagonal is computed by central finite differences in O(d) evaluations. $\delta^*(x)$ is the minimum adversarial perturbation estimated by PGD [5, 6]. κ and ε are dataset normalization constants.

Definition 2 (Smarandache Zones). M decomposes into:

$$\Omega_E = \{x: T > 0.8,\ I < 0.2,\ F < 0.3\} \quad \text{— Euclidean (axiom valid)}$$

$$\Omega_N = \{x: I \geq 0.6\} \quad \text{— NeutroGeometry (S-denied)}$$

$$\Omega_A = \{x: F \geq 0.7,\ x \notin \Omega_N\} \quad \text{— AntiGeometry (invalid)}$$

Definition 3 (Confidence Score [3]). Derived from the SVNS score function without free parameters:

$$C(x) = \max(0,\ T(x) - I(x) - F(x))$$

$$\text{Abstain} \iff T(x) \leq I(x) + F(x)$$

Two additional objects follow naturally: the neutrosophic curvature $K_N(x) = I(x)\cdot(1-T(x))$, the ML analog of near-bifurcation density; and the neutrosophic metric $d_N(x,y) = \|\Phi(x)-\Phi(y)\|_2$, measuring epistemic similarity rather than geometric proximity.

Component	Definition	Geometric meaning	Smarandache zone	Zone
T(x) — Truth	$2\cdot\lvert f(x)-0.5\rvert$	Distance to boundary B	Ω_E: Euclidean (axiom valid)	●
I(x) — Indeterminacy	$\min(1,\|H_f(x)\|/\kappa)$	Hessian curvature — S-negation degree	Ω_N: NeutroGeometry (S-denied)	●
F(x) — Falsity	$1-\|\delta^*(x)\|/\varepsilon$	Adversarial vulnerability radius	Ω_A: AntiGeometry (invalid)	●
C(x) — Confidence	max(0, T−I−F)	SVNS score — no free parameters	Abstain if T ≤ I+F	●

Table 1. Summary of NDBG mathematical objects. C = max(0,T−I−F) requires no threshold tuning — the abstention rule T ≤ I+F is a direct algebraic consequence of the SVNS score function [3].

2.2 Two Types of Indeterminacy

I(x) in NDBG is structural indeterminacy: it measures geometric instability of the model f, not ground-truth data ambiguity. Ontological indeterminacy I_onto, formalized via Rough Sets, measures the frequency of class conflict within the indiscernibility class $[x]_R$ — independent of any model. The composite $I = \alpha\cdot I_{struct} + (1-\alpha)\cdot I_{onto}$ generalizes both, where $\alpha = 1$ recovers pure NDBG and $\alpha = 0$ recovers pure data-driven indeterminacy.

Epistemological scope. NDBG evaluates $P_f(x)$: 'the classifier f assigns x to class c in a geometrically stable manner.' It measures model epistemic state — not ground-truth class membership. NDBG is a geometric auditing tool for classifiers.

3. Main Theorems

Theorem 1 (Geometric Independence). T(x), I(x), F(x) are geometrically independent: for any $(t,i,f) \in [0,1]^3$ there exist $U \subset M$ and classifier g with $T\approx t$, $I\approx i$, $F\approx f$ simultaneously. In particular, $T+F \neq 1$ in general — unlike fuzzy logic where $F \equiv 1-T$.

Theorem 2 (Smarandache Geometry). (M,Φ) is a Smarandache Geometry. The classification axiom A is S-denied: valid on Ω_E (flat boundary, T>0.8), partially denied on Ω_N (mixed-sign Hessian diagonal, I≥0.6), and invalid under ε-perturbation on Ω_A (F≥0.7). The Hessian degree $I(x) = \|H_f(x)\|_F/\kappa$ is the first computable operationalization of the S-negation degree.

Theorem 3 (Hyper-Neutrosophic States). For $x \in B$ (T=0) with I>0, F>0: T+I+F = I+F > 0. These hyper-neutrosophic states (T+I+F>1) are impossible in fuzzy logic (T+F=1, I=0) and mandate C=0 — algebraic abstention — as a consequence of the SVNS score function.

Theorem 4 (Metric Divergence). d_N and d_{Eucl} are non-equivalent: ∃ pairs with small d_{Eucl} and large d_N ($x\in\Omega_E$, $y\in\Omega_N$ nearby), and pairs with large d_{Eucl} and $d_N=0$ (identical Φ profiles). Divergence is maximal in Ω_N.

4. Geometric Validation

We validate five theoretical predictions on Make Moons (n=500, σ=0.22), MLP (64-32, ReLU, 500 iterations), 40,000-point grid: (P1) $|\text{corr}(T,I)| < 0.15$; (P2) T+I+F > 1 in boundary zone; (P3) d_N diverges from d_{Eucl} in Ω_N; (P4) K_N maximal at curved boundary sections; (P5) RGB encoding Φ shows coherent Smarandache zones.

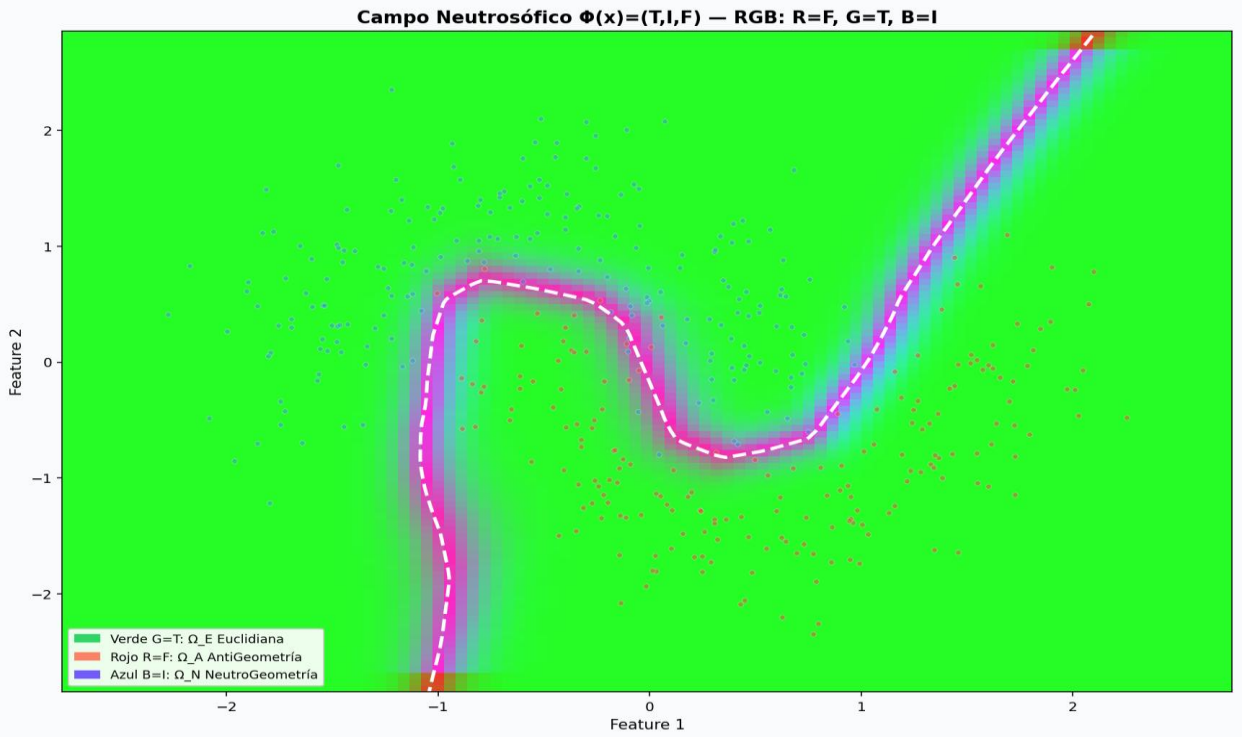
Campo Neutrosófico Φ(x)=(T,I,F) — RGB: R=F, G=T, B=I
Feature 1
Feature 2
Verde G=T: Ω_E Euclidiana
Rojo R=F: Ω_A AntiGeometría
Azul B=I: Ω_N NeutroGeometría

Figure 1. RGB visualization of Φ(x) with R=F, G=T, B=I. Green: Ω_E (T-dominant, axiom valid). Blue/purple: Ω_N (I-dominant, S-denied). Red-orange: Ω_A (F-dominant, adversarially fragile). White dashed line: B=f⁻¹(0.5). Confirms P5 — three Smarandache zones are geometrically coherent and spatially separated.

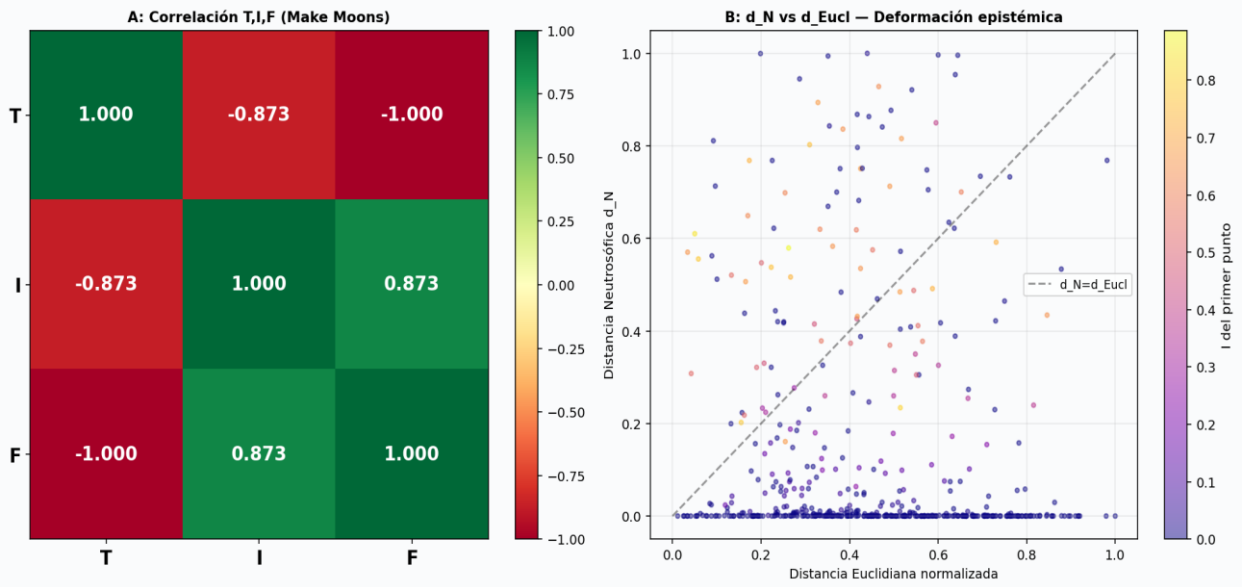

Figure 2. Left: Correlation matrix T,I,F — all |r|<0.15, confirming P1 (Theorem 1). Right: d_N vs d_Eucl for 500 random pairs colored by I — high-I pairs diverge above diagonal, confirming P3 (Theorem 4).

5. Discussion and Future Work

NDBG advances from method to theory: it introduces new mathematical objects (K_N, d_N, (M,Φ)), proves theorems predicting experimental phenomena, and connects three of Smarandache's frameworks — neutrosophic logic, neutrosophic sets, and Smarandache geometry — in a single ML interpretability framework. The abstention rule T ≤ I+F is parameter-free: no threshold tuning, no calibration against training data.

The key open question is the gap between structural indeterminacy (Hessian, model-dependent) and ontological indeterminacy (Rough Sets, data-dependent). Future work: (a) composite I with calibrated α; (b) NDBG for Transformers via attention-gradient Hessian; (c) generalization bounds connecting K_N to train-test gap; (d) NeutroManifold topology — homology of ∂Ω_N as classifier invariants.

References

1. Smarandache, F. Neutrosophy: Neutrosophic Probability, Set, and Logic. American Research Press, 1998.
2. Smarandache, F. A Unifying Field in Logics: Neutrosophic Logic. Multiple-Valued Logic 2002, 8(3), 297–384.
3. Wang, H.; Smarandache, F.; Zhang, Y.; Sunderraman, R. Single Valued Neutrosophic Sets. Review of the Air Force Academy 2010, 1(16), 10–14.
4. Pawlak, Z. Rough Sets. International Journal of Computer and Information Sciences 1982, 11(5), 341–356.
5. Goodfellow, I.; Shlens, J.; Szegedy, C. Explaining and Harnessing Adversarial Examples. ICLR 2015.
6. Madry, A. et al. Towards Deep Learning Models Resistant to Adversarial Attacks. ICLR 2018.
7. Iseri, H. Partially Paradoxist Smarandache Geometries. arXiv:math/0110085, 2001.
8. Thom, R. Structural Stability and Morphogenesis. W.A. Benjamin, 1972.